

Table: Summary of techniques used across top 10 phenotypes.

Phenotype	Rule Based	NLP	Data Mining	Hybrid	System Description
Asthma	1	2	1	1	0
Cancer	5	10	14	6	2
Cataract		2	1		1
Diabetes	7	9	5	2	7
Drug Side Effect		1	3	1	
Heart Failure		2	4		
Hypertension		2			1
Peripheral Arterial Disease		1	1		1
Pneumonia		4	3	3	1
Rheumatoid Arthritis		4	3	3	

Note: A study that uses a combination of rules and NLP is termed as hybrid. The above table counts such a study in each of the three columns, namely, rule based, NLP and hybrid. System descriptions include studies addressing the issue of cohort identification from a system point of view. For example: architecture description of a large scale system or best practice, evaluation of a data model.